The Data Warehouse

A Data Warehouse is a method of making the accumulated data of a business available to decision-makers. It can help managers to spot trends, help to forecast the future of the business, and aid in setting corporate strategy.

Operational and Analytical Data

In any business organization there are two types of data: operational data and analytical data. Operational data consists of information which is tactical. It is data which is needed to run the day-to-day operations of a business. Examples include invoicing, account balances and inventory control. It is critical to keep operational data accurate and current. When a retail customer pays a bill or a banking customer deposits a check, the database must be immediately updated.

Analytical data is strategic information. It helps to identify trends, recognize opportunities and pinpoint problems. It is data about data and is used for business decision making. Analytical data is what Data Warehouses are all about.

Analytical data must be sheltered from the constant updating and modification of operational data. Analytical data is not accurate to the moment, but is accurate over a specific period of time which could be a week, month, or year.

You may have heard the saying “Can not see the forest for the trees.” In this case Analytical data sees the forest; Operational data see the trees.

Ultimately, operational and analytical data are too different to be kept together. A good analytical database makes a bad operational database and a good operational database makes a bad analytical database. In an operational database there are many transactions which affect small portions of data. Operational database utilization is predictable and the system can be optimized for projected workloads. Analytical database usage is less predictable. It processes a relatively small number of queries but accesses large amounts of data. The result is that the databases cannot be mixed. They must be separately maintained.

Extract Programs

Data Warehouses get their data from operational databases. Periodically an extract program is run which moves detailed transaction data to the data warehouse. This information is added to the data from previous extractions to give the volume of data that management needs to make business decisions.

Asking the Right Questions

Perhaps one of the most important issues in regard to Data Warehousing is the importance of asking the right questions. A Data Warehouse can be queried in an infinite number of ways, but

The quality of the extracted information is a direct result of the quality of the queries. In the case of Data Warehouses the end user must be a highly knowledgeable business analyst to fully benefit from the Data Warehouse capabilities.

Another challenge is deciphering the information once you have it. What is the information telling us and what importance should be placed on new and unlikely connections between the data? Are we too smart for our own good? Will we fool ourselves? What is the data telling us if anything at all?

Data Warehouse Characteristics

Regardless of the make, model, or vendor, all good Data Warehouses will have the following characteristics:

- Separate from the operational database.
- Based on a comprehensive model of the organization information resources.
- Data delivered from multiple operational databases.
- Automatically maintained.

- Can be accessed by many query tools.
- Is read-only.
- Can be replicated. No need to keep information in a central location.
- Information is presented in consistent and compatible formats. Example - Revenue is presented and converted to one currency such as U.S. Dollars or German Marks.

Data Warehouse Terminology

Here are a few terms that are often used when discussing Data Warehousing concepts.

Datamart - A subset of a Data Warehouse for a single department or function. Is usually much smaller than a data warehouse.

Decision Support System - (DSS) Extract and report software programs used by management to help make strategic decisions. Usually looks at a specific type of data to provide reports. May or may not be used in conjunction with a Data Warehouse. Provides the ability to model by answering "what if" types of questions.

Executive Information System - (EIS) Provides a concise snapshot of how a business is doing on a daily basis. Should be presented on one screen if possible. Often uses graphical elements to transfer knowledge.

Key Points to Remember

- Business oriented databases are composed of two types of data: Operational Data and Analytical Data.
- A Data Warehouse is populated with Analytical Data.
- Data Warehouses are used for strategic decision making.
- Analytical Data must be managed and maintained separately
- From Operational Data.
- Data in a Data Warehouse is "read-only" and cannot be altered by the end user.
- Analytical data gives a "snapshot" of a period of time and is not up-to-the-moment current.